

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO – MATEMATIČKI FAKULTET
BIOLOŠKI ODSJEK

LJUDSKA GENETIČKA VARIJABILNOST

HUMAN GENETIC VARIATION

SEMINARSKI RAD

Hrvoje Mišetić
Preddiplomski studij Molekularna biologija
Undergraduate Study of Molecular Biology
Mentor: doc. dr. sc. Damjan Franjević

Zagreb, 2016.

SADRŽAJ

1. UVOD	2
2. OBLICI VARIJABILNOSTI.....	3
2.1. Pojedinačni nukleotidni polimorfizmi	4
2.2. Strukturalne varijante	5
3. ISTRAŽIVANJE LJUDSKE GENETIČKE VARIJABILNOSTI	6
3.1. The International HapMap Project	6
3.1.1. Istraživanje	7
3.1.2. Rezultati	8
3.2. The 1000 Genomes Project	9
3.2.1. Ciljevi	9
3.2.2. Pilot faza	9
3.2.3. Prva faza	12
3.2.4. Izrada mape varijabilnosti	13
3.2.5. Druga faza	15
3.2.6. Treća faza	15
4. EVOLCIJA I GENETIČKA VARIJABILNOST	16
4.1. Evolucija Hominida	17
4.1.1. Genetička varijabilnost Hominida	19
4.2. Evolucija čovjeka	20
4.2.1. Populacijska genetika	22
4.2.1. Križanje modernog čovjeka i arhaičnih populacija	23
5. GENETIČKA VARIJABILNOST I RASE.....	24
6. SAŽETAK	25
7. LITERATURA.....	26

1. UVOD

U posljednja dva desetljeća svjedoci smo naglog razvoja tehnologija sekvenciranja genoma i metoda njegove analize. Konstantno usavršavanje točnosti i brzine sekvenciranja uz postepeni pad cijene omogućili su masovno korištenje ove tehnologije i stvaranje enormne količine podataka o ljudskom genomu. Novonastali podaci omogućili su istraživanje i analizu ljudskog genoma te posebice proučavanja razlika u genomu između ljudi i različitih populacija, tj. ljudske genetičke varijabilnosti.

Ljudska genetička varijabilnost predstavlja genetičke razlike unutar i između populacija ljudi. Prosječno, na razini slijeda nukleotida u molekuli DNA svi ljudi dijele 99.5 % sličnosti (Levy S. et al. , 2007.).

Uzroci genetičkih razlika između jedinki su neovisna segregacija kromosoma i homologna rekombinacija tijekom mejoze i različite vrste mutacija koje se pojavljuju tijekom razvoja i čitavog života jedinke. Postoje barem dva razloga koja mogu objasniti zašto postoji genetička varijabilnost između populacija. Prirodna selekcija može doprinijeti adaptivnoj prednosti jedinke u određenom okolišu ako određeni alel omogućuje kompetitivnu prednost. Aleli pod utjecajem selekcije će se vjerojatnije pojaviti u onim geografskim područjima gdje omogućuju prednost jedinki i preživljavanje. Drugi razlog postojanja genetičke varijabilnosti je visoka razina neutralnosti većine genetičkih mutacija. Većina mutacija nema nikakav selektivni efekt na organizam, glavni uzrok tome je genetički drift - promjene u zalihi gena izazvane slučajnošću. Kod ljudi, efekt osnivača i nekadašnja mala populacija u prošlosti koja povećava vjerojatnost genetičkog drifta vrlo vjerojatno je imala veliki utjecaj na genetičke razlike između populacija ljudi.

Otkrića u istraživanju ljudske genetičke varijabilnosti omogućila su razumijevanje naših razlika i sličnosti, našeg podrijetla i rekonstruiranje evolucijske povijesti čovjeka. Pomažu znanstvenicima razumjeti i otkriti migracije različitih populacija ljudi tijekom ljudske povijesti te odrediti srodstvene i biološke odnose između tih populacija. Uz doprinos evolucijskim istraživanjima, otkrića iz područja ljudske genetičke varijabilnosti imaju veliku primjenu u medicini. Nastoje se povezati promjene u genotipu s promjenama u fenotipu pogotovo one promjene koje uzrokuju kompleksne poligenske bolesti poput dijabetesa, srčanih bolesti i raka. Poznavajući genetičke uzroke tih bolesti nastoji se kreirati najbolja terapija za bolesnika te omogućiti personalizirana medicina u pravom smislu te riječi.

2. OBLICI VARIJABILNOSTI

Genetička varijabilnost u ljudi pojavljuje se na više razina, od većih promjena na razini kariotipa do mutacije samo određenog nukleotida. Nukleotidna različitost (engl. nucleotide diversity) je prosječan broj nukleotida koji razlikuje dvije jedinke. Ljudska nukleotidna različitost je procijenjena na između 0.1% i 0.4% od ukupnog broja parova baza (Jorde, L. and Wooding, S., 2004.). To predstavlja razliku između 3 do 12 milijuna parova baza budući da ljudski genom ima oko 3 milijardi baza u haploidnoj stanici. Prema podacima The 1000 Genomes Project-a genom pojedinog čovjeka razlikuje se od referentnog ljudskog genoma na 4.1 do 5.0 milijuna mjesta i obuhvaća 20 milijuna parova baza u slijedu DNA (The 1000 Genomes Project Consortium, 2015.). Na tim mjestima mogu se pojavljivati pojedinačni nukleotidni polimorfizmi (engl. single nucleotide polymorphisms), kratke insercije odnosno delecije, promjene broja ponavljanja (engl. copy number variant) koje sve osim pojedinačnih nukleotidnih polimorfizama pripadaju u grupu strukturalnih varijanata (engl. structural variants). Ne postoji dogovorena nomenklatura koja opisuje različite razrede ljudskih genetičkih varijanata.



Slika 1. Terminologija koja se koristi u opisivanju ljudske genetičke varijabilnosti

(Frazer, K., Murray, S., Schork, N. and Topol, E., 2009.)

Međutim, postoji terminologija koja opisuje različite vrste varijanata prema njihovom specifičnom nukleotidnom rasporedu koji ih razlikuje od drugih tipova varijanata.

Prema Slici 1. to su varijante pojedinačnog nukleotida (engl. single nucleotide variant), insercija-delecija varijanta (engl. insertion-deletion variant), blokove supstitucije (engl. block substitution), inverzijski varijant (engl. inversion variant), promjena broja kopija ponavljanja (engl. copy number variant).

Pojedinačni nukleotidni varijanti slijed u DNA je slijed u kojem je samo jedan nukleotid (A, T, G ili C) promijenjen.

Insercija-delecija varijant pojavljuje se kada su prisutne jedna ili više dušičnih baza u nekom genomu, tj. odsutne u nekom drugom genomu. Najčešće se sastoje od nekoliko baza, ali mogu biti čak veći od 80 kb.

Blokovske supstitucije opisuju slučajeve u kojima se niz susjednih nukleotida razlikuje između dva genoma.

Inverzijski varijant je onaj u kojem je raspored baznih parova je obrnut u određenom dijelu kromosoma.

Različit broj kopija ponavljanja se pojavljuje kada se identičan ili gotovo identičan slijed ponavlja u određenom dijelu kromosoma, a drugom genomu je broj ponavljanja manji ili veći.

2.1. POJEDINAČNI NUKLEOTIDNI POLIMORFIZMI

Pojedinačni nukleotidni polimorfizam (engl. single nucleotide polymorphisms) je razlika u jednom nukleotidu određenog slijeda DNA između jedinki jedne vrste koja se pojavljuje s učestalošću od barem 1 % u populaciji. Procjenjuje se da ljudski genom sadrži između 10 do 30 milijuna pojedinačnih nukleotidnih polimorfizama.

To je najčešći tip varijabilnosti između dva slijeda DNA i obuhvaća 90 % ukupnog broja varijabilnosti između dva slijeda odnosno genoma (Collins, F. S.; Brooks, L. D.; Chakravarti, A., 1998.). Dije se u dvije grupe prema tome kakav utjecaj imaju na fenotip jedinke.

Funkcionalne ili nesinonimne točkaste mutacije su one koje utječu na izrezivanje gena ili na mRNA te time uzrokuju razlike u fenotipu jedinki određene vrste.

Kod čovjeka između 3% i 5% svih točkastih mutacija su funkcionalne (International HapMap Consortium, 2010.). Neutralne ili sinonimne točkaste mutacije su one koje ne uzrokuju promjenu u izrezivanju gena i ne utječe na mRNA te time ne uzrokuje promjenu fenotipa.

Najčešće se koriste kao genetički markeri u istraživanjima zbog stabilnog nasljeđivanja tijekom niza generacija.

Postoje i kodirajuće točkaste mutacije koje se pojavljuju unutar gena. Zasad je kod čovjeka popisano 105 takvih mutacija koje uzrokuju prerane stop kodone u 103 gena. Takva vrsta

točkastih mutacija čini 0.5 % zasada otkrivenih kodirajućih točkastih mutacija. Pojavljuju se zbog djelomične duplikacije u genomu i uzrokuju gubitak funkcije proteina (Ng et al., 2008.) Većina točkastih mutacija je popisana i identificirana te informacije o njima nalaze se u The Single Nucleotide Polymorphism database (dbSNP).

2.2. STRUKTURALNE VARIJANTE

U strukturalne varijante (engl. structural variants) pripadaju sve genetičke varijante koje nisu točkaste mutacije. To uključuje insercije, delecije, inverzije, duplikacije i različit broj kopija ponavljanja. Najčešće obuhvaćaju promjene od 1 Kb do 3 Mb što je veće od točkaste mutacije, a manje od kromosomske abnormalnosti. Prema rezultatima The 1000 Genomes Project-a, tipičan ljudski genom sadrži od 2,100 do 2,500 strukturalnih varijanata (oko 1,000 većih delecija, oko 160 različitih broja kopija ponavljanja, oko 915 Alu insercija, oko 128 L1 insercija, oko 51 SVA insercija, oko 4 NUMTs i oko 10 inverzija) koje obuhvaćaju oko 20 milijuna baza (Auton et al., 2015).

Jedno od glavnih pitanja je može li se na temelju položaja točkastih mutacija u genomu predvidjeti položaj strukturalnih varijanti, tj. postoji li između njih neravnoteža udruživanja (engl. linkage disequilibrium) - neslučajna povezanost alela na različitim lokusima. Više različitih istraživanja je pokazalo da česti kratki indeli kao i veći strukturalni polimorfizmi u jedinstvenim regijama genoma su u neravnoteži udruživanja s označenim točkastim mutacijama. Strukturalne varijante se naime slično ponašaju kao točkaste mutacije u pogledu genomske i populacijske distribucije, navodeći sličnu evolucijsku povijest. Oba tipa varijanata su ancestralni, pojavivši se jednom u ljudskoj povijesti su se nastavili nasljeđivati među generacijama te su se u puno manjoj mjeri pojavljivale kao rezultat povratnih mutacija (Frazer, K., Murray, S., Schork, N. and Topol, E., 2009.).

3. ISTRAŽIVANJE LJUDSKE GENETIČKE VARIJABILNOSTI

2001. godine u sklopu Human Genome Project-a prvi put u povijesti čovječanstva je sekvenciran genom čovjeka. Od tog trenutka koji je obilježio povijest i pokrenuo novu eru u istraživanju genetike čovjeka, tehnologije sekvenciranja su se usavršavale i visoke cijene sekvenciranja su padale te se time omogućilo pokretanje istraživanja koja su bavili onim što čini ljude različitim, tj. genetičkom varijabilnošću.

Dva projekta koja se izdvajaju od ostalih su The International HapMap Project i The 1000 Genomes Project. Glavni cilj oba projekta je bio otkriti, popisati i analizirati sve oblike genetičke varijabilnosti čovjeka. Kreirale su se baze podataka sa svim informacijama vezanim za ljudsku genetičku varijabilnost i te baze su javno dostupne znanstvenoj zajednici.

Daljnji cilj istraživanja je različite varijante genotipa povezati s promjenama u fenotipu, tj. povezati promjene genotipa koje se u određenoj frekvenciji pojavljuju u određenim populacijama s kompleksnim poligenskim bolestima.

O veličini i važnosti ovih projekata govori podatak da je prva publikacija The 1000 Genomes Project-a jedan od najcitiranijih radova u biologiji uopće.

3.1. The International HapMap Project

The International HapMap Project je projekt čiji cilj je bio napraviti haplotipsku kartu ljudskog genoma, opisati opći uzorak ljudske genetičke varijabilnosti.

Ovaj projekt je nastao suradnjom znanstvenika iz akademske zajednice, neprofitabilnih biomedicinskih istraživačkih grupa i privatnih kompanija iz Kanade, Kine, Ujedinjenog Kraljevstva, Japana, Nigerije i Sjedinjenih Američkih Država. Započeo je službeno sastankom u listopadu 2002. godine, a završio na proljeće 2009. godine. U tom periodu podaci o varijabilnosti i analize istih izlazi su tri puta 2005., 2007. i 2009. godine.

3.1.1 Istraživanje

Glavni cilj je bio povezati razlike u ljudskim genomima s kompleksnim poligenским bolestima koje nastaju međusobnim djelovanjem više gena i okolišnih faktora te su zbog toga puno teže za istraživanje od monogenских bolesti koje su k tome i puno rjeđe.

U poligenске bolesti pripadaju dijabetes, rak, srčane bolesti te razni psihološki poremećaji. Najlogičniji pristup kojim bi se povezao genotip s nekom od navedenih bolesti je sekvenciranja genoma zdravih jedinki i bolesnih jedinki te potom tražiti razlike između dva seta genoma međutim taj pristup tada nije bio moguć zbog visoke cijene sekvenciranja čitavog genoma.

Budući da ljudi dijele oko 99.5% DNA slijeda u sklopu ovog projekta genotipizirani su samo najčešći pojedinačni nukleotidni polimorfizmi od 269 jedinki i objavljene u javno dostupnoj bazi podataka.

Aleli obližnjih pojedinačnih nukleotidnih polimorfizama na pojedinom kromosomu su u korelaciji. Ako je alel jednog pojedinačnog nukleotidnog polimorfizma u nekoj jedinci poznat, onda aleli obližnjih pojedinačnih nukleotidnih polimorfizama se često mogu predvidjeti. Svaki pojedinačni nukleotidni polimorfizam se pojavio u evolucijskoj povijesti kao točkasta mutacija na slijedu DNA okružen drugim, ranijim točkastim mutacijama. Pojedinačni nukleotidni polimorfizmi koji su vrlo udaljeni na kromosomu nisu u velikoj korelaciji zbog rekombinacije koja se pojavljuje u svakoj generaciji i miješa alele dva kromosoma. Slijed uzastopnih alela na određenom kromosomu naziva se haplotip.

Kako bi se otkrili genetički faktori odgovorni za pojedine bolesti, uzme se određeni dio kromosoma za koji je poznato da se na njemu nalaze mogući geni za određenu bolest. Na temelju pojedinačnih nukleotidnih polimorfizama koji su otkriveni u projektu i njihove međusobne korelacije određuje se haplotip. Usporedbom genotipa bolesnih i zdravih jedinki saznaju se točne lokacije i haplotipovi koji su uzročnici bolesti.

3.1.2. Rezultati

Aleli pojedinačnih nukleotidnih polimorfizama locirani u istom genomskom intervalu su često u korelaciji jedan s obzirom na drugi. Neravnoteža udruživanja varira na kompleksan i nepredvidljiv način unutar genoma i između različitih populacija. Trud uložen u prvoj fazi projekta popločio je put rastavljanju genoma na grupe visoko koreliranih pojedinačnih nukleotidnih polimorfizama koji se zajedno nasljeđuju (engl. LD bins). U drugoj fazi projekta utvrđeno je da velika većina pojedinačnih nukleotidnih polimorfizama s niskom alelnom frekvencijom (engl. minor allele frequency) od barem 5 % mogu biti reducirani do ~550,000 grupa koje se nasljeđuju skupa u jedinkama europskog ili azijskog podrijetla i do 1,100,000 grupa za jedinke afričkog podrijetla. U posljednjoj trećoj fazi genotipizirano je 1,6 milijuna pojedinačnih nukleotidnih polimorfizama iz 1,184 jedinke iz 11 populacija. Tim je proširen javni resurs varijanata iz globalnih populacija i napravljen prvi veći korak prema stvaranju visoko rezolucijske karte ljudske genetičke varijabilnosti (Altshuler et al., 2010).

3.2. The 1000 Genomes Project

The 1000 Genomes Project trajao je od 2008. do 2015. godine stvorivši najveći javno dostupan katalog ljudske genetičke varijabilnosti do sada.

Cilj projekta je bio pronaći većinu genetičkih varijanata s frekvencijom od barem 1% u proučavanim populacijama. To je bio prvi projekt u kojem je sekvenciran genom velikog broja ljudi i svi podaci kreirani u projektu su javno dostupni putem besplatnih baza podataka. Iako je sekvenciranje i dalje bilo preskupo da bi se svi genomi sekvencirali s najvećom točnošću korištenjem najnovijih tehnologija omogućen je dobar omjer cijene i kvalitete.

Projekt je bio proveden u četiri faze - pilot fazi i tri faze glavnog projekta. U prvoj i trećoj fazi glavnog projekta stvarali su se podaci dok se u drugoj fazi bavilo tehničkim razvojem.

3.2.1. Ciljevi

Glavni ciljevi projekta su bili:

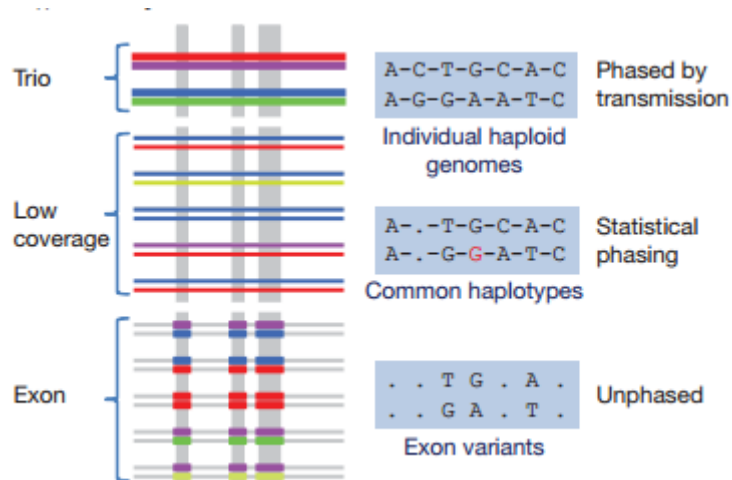
- pronaći 95 % svih pojedinačnih nukleotidnih polimorfizama s učestalošću od barem 1% u istraživanim populacijama (u kodirajućim regijama genoma ciljalo se na 0.1 %)
- pronaći kratke indele i strukturalne varijante
- odrediti genotip i haplotip svih uzoraka
- objavljivati podatke javno i brzo
- razviti metode analize sljedova DNA, alate i resurse koji se kasnije mogu koristiti u drugim projektima

3.2.2. Pilot faza

Pilot faza projekta je provedena kako bi se na temelju njenih rezultata najbolje dizajnirao i organizirao glavni projekt. U sklopu pilot faze provedena su tri manja projekta:

1. Sekvenciranje čitavog genoma niskom pokrivenošću (engl. low coverage) 179 jedinki iz 4 različite populacije
2. Sekvenciranje čitavog genoma visokom pokrivenošću (engl. high coverage) dva tripleta majka-otac-dijete
3. Sekvenciranje eksona 697 jedinki iz sedam populacija

Na temelju podataka dobivenih iz ova tri manja projekta trebalo je odrediti hoće li sekvenciranje niskom pokrivenošću zadovoljiti ciljeve projekta.



Slika 2. Grafički prikaz rezultata Pilot faze (Durbin et al., 2010)

Na Slici 2. boje na lijevoj strani označuju različite haplotipove u pojedinim genomima, a širina linija predstavlja dubinu pokrivenosti sekvenciranja. Osjenčano područje na desnoj strani daje primjer rezultata koji mogu biti dobiveni trima različitim strategijama korištenim u pilot fazi (točke predstavljaju podatke koji nedostaju, crtice predstavljaju jesu li sljedovi separirani s obzirom potječu li od oca ili majke (engl. phase information), tj. jesu li heterozigotne varijante povezane s točnim haplotipom). Unutar kratke regije genoma svaka jedinka ima dva haplotipa koja dijeli s nekim u populaciji.

1. U triplet projektu korišteno je sekvenciranje visoke pokrivenosti i omogućilo je precizno otkrivanje velikog broja varijanata u čitavom genomu. Genotipizaciji, određivanju haplotipa i

provjeri kvalitete je pomoglo Mendelovo nasljeđivanje budući da su uzorci tripleti majka otac dijete.

2. Sekvenciranje niske pokrivenosti uspješno je odredilo zajedničke varijante na poznatim haplotipovima, ali je imalo manju moć detekcije rjeđih haplotipova te daje nekada netočne genotipove.
3. Sekvenciranje eksona omogućilo je poprilično točno otkrivanje čestih, i rijetkih varijanata, ali samo u tim ciljanim područjima jer izvan njih nemaju sposobnost odrediti varijante (Durbin et al., 2010).

Sva tri projekta su dijelila sljedeća četiri koraka u određivanju genetičkih varijanata:

1. Otkrivanje varijanata: sravnjenje sekvenciranih dijelova genoma s referentnim genomom i identifikacija mjesta i regija na kojima bi se genom jedinki razlikovao od referentnog slijeda. Određivanje varijanata je bilo ograničeno na pristupačan genom (engl. accessible genome). To je dio referentnog genoma koji preostane nakon što se uklone oni dijelovi s mnogo dvosmisleno smještenih očitavanja (engl. reads) ili s neočekivano puno ili malo sravnjenih očitavanja.
2. Filtriranje: korištenjem kontrola kvalitete uklanjali su se rezultati koji su vjerojatno lažni pozitivni.
3. Genotipizacija: procjena alela prisutnih u svakoj jedinki na mjestima ili dijelovima koji su varijabilni.
4. Procjena točnosti: korištenjem neovisnih tehnologija na novootkrivenim genetičkim varijantama određivala se razina netočnih rezultata (engl. false discovery rate)

Na kvalitetu novootkrivenih varijanata utječu mnogi faktori uključujući točnost sekvenciranja, pouzdanost sravnjenja s referentnim genomom i metode koje određuju individualni genotip.

Rezultati pilot faze nalažu da sekvenciranje niske pokrivenosti pruža efikasan pristup otkrivanju varijabilnosti na razini genoma, a sekvenciranje eksona omogućuje efikasno određivanje i točno genotipiziranje rjeđih varijanata u regijama koje su od većeg interesa.

Na temelju projekta s dva para tripleta procijenjeno je da je stopa de novo mutacija u spolnim embrionalnim prastanicama 10^{-8} mutacija po bazi po generaciji (Durbin et al., 2010).

Otkriveno je više 95% općih varijanata koji su bili tada poznati u nekoj jedinci, ali oni rjeđi varijanti posebice izvan eksona ostali su slabo opisani.

3.2.3. Prva faza

U prvoj fazi genomi 1,092 jedinke iz 14 populacija koje potječu iz Europe, Istočne Azije, subsaharskog područja Afrike te Sjeverne i Južne Amerike su sekvencirani i analizirani korištenjem sekvenciranja niske pokrivenosti (2-6X) i ciljanog sekvenciranja eksona (engl. targeted deep exome sequencing) (50-100X). Ovaj odabir tehnologija je pokazao se najisplativijim i najposobnijim za otkrivanje genetičke varijabilnosti. Korištenjem statističkih metoda za odabir genetičkih varijanata više kvalitete od mogućih kandidata dobivenih multiplim algoritmima i integracijom pojedinačnih nukleotidnih polimorfizama, indela i većih strukturalnih varijanata u jednu cjelinu, dodatno je proširen čitav pristup.

Tablica 1. Rezultati prve faze projekta (McVean et al., 2012)

	Autosomes	Chromosome X	GENCODE regions*
Samples	1,092	1,092	1,092
Total raw bases (Gb)	19,049	804	327
Mean mapped depth (x)	5.1	3.9	80.3
SNPs			
No. sites overall	36.7 M	1.3 M	498 K
Novelty rate†	58%	77%	50%
No. synonymous/non-synonymous/nonsense	NA	4.7/6.5/0.097 K	199/293/6.3 K
Average no. SNPs per sample	3.60 M	105 K	24.0 K
Indels			
No. sites overall	1.38 M	59 K	1,867
Novelty rate†	62%	73%	54%
No. inframe/frameshift	NA	19/14	719/1,066
Average no. indels per sample	344 K	13 K	440
Genotyped large deletions			
No. sites overall	13.8 K	432	847
Novelty rate†	54%	54%	50%
Average no. variants per sample	717	26	39

NA, not applicable.

* Autosomal genes only.

† Compared with dbSNP release 135 (Oct 2011), excluding contribution from phase I 1000 Genomes Project (or equivalent data for large deletions).

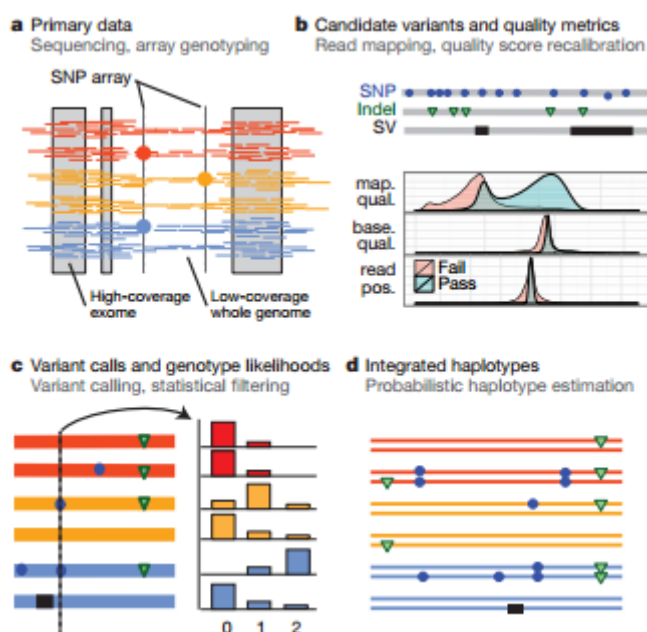
Prema podacima iz Tablice 1. u prvoj fazi projekta otkriveno je ukupno 38 milijuna pojedinačnih nukleotidnih polimorfizama, 1.4 milijuna kratkih insercija i delecija te više od 14 000 većih delecija.

Najvažniji zaključci nakon prve faze projekta su bili sljedeći:

1. Jedinke iz različitih populacija imaju različite profile rijetkih i čestih varijanata te varijante niske frekvencije imaju znatnu geografsku diferencijaciju.
2. Varijante niske frekvencije su bile loše opisane u pilot fazi posebno one van kodirajućih regija, a one su od iznimne važnosti jer su bogate potencijalno funkcionalnim mutacijama za na primjer, varijantama koje uzrokuju promjene u funkciji proteina.
3. Nadalje varijante niske frekvencije su najčešće recentnog nastanka i prikazuju povećanu razinu diferencijacije populacije.

4. Određivanje varijanata niske frekvencije u populacijama pomaže identificirati mnogo varijanata koji imaju funkcionalnu važnost i neizbježni su za interpretaciju individualnih genoma, i odvajanje zajedničkih varijanata od onih koji su karakteristični za manje grupe.

3.2.4. Izrada integrirane mape varijabilnosti



Slika 3. Čitav proces izrade mape varijabilnosti od dobivanja sljedova sekvencirane DNA do samog otkrića genetičkih varijanata (McVean et al., 2012)

Prema Slici 3. postoje četiri glavna koraka u određivanju genetičke varijabilnosti nekog slijeda DNA:

A Primarni podaci, tj. sekvencirani genom svake jedinke nastao je sekvenciranjem čitavog genoma niskom pokrivenošću (prosječna pokrivenost 5X) i visokom pokrivenošću (prosječna pokrivenost 80X) su određeni sljedovi eksona i korišten je (engl. high density SNP array).

B Sravnjenjem očitavanja (engl. read) s referentnim genomom te korištenjem višestrukih algoritama dobivaju se moguće genetičke varijante. Za svaku varijantu, određene su mjere kvalitete (engl. quality metrics) koje uključuju informacije o jedinstvenosti okružujuće sekvence, kvaliteta dokaza koja potkrepljuju varijantu (npr. kvaliteta baza (engl. base quality) i pozicija baza varijanta unutar očitavanja), i distribucija varijanata u populaciji. Metode strojnog učenja koriste ove multidimenzionalne informacije kako bi poredali varijante po

pouzdanosti i odredili naknadno prag točnosti kako bi se osigurala niska stopa netočnih rezultata (engl. false-discovery rate)

C Vjerojatnosti genotipa (engl. genotype likelihoods) su korišteni kako bi saželi dokaze za svaki genotip na bialelnim mjestima (0, 1 ili 2 kopije varijanta) u svakom uzorku na svakom mjestu.

D Kako je dokaz za svaki genotip tipično slab u podacima dobivenim niskom pokrivenošću i mogu biti visoko varijabilni s podacima dobivenim sekvenciranjem eksona, statističke metode su korištene kako bi ujednačile informacije iz uzoraka povezanosti ravnoteže (engl. linkage disequilibrium), omogućujući što točnije određivanje haplotipa i genotipa.

Uspjeh sekvenciranja eksoma u otkrivanju bolesti koje se nasljeđuju po Mendelu i otkriće rijetkih i nisko frekventnih varijanata povezanih s bolestima u genima asociranim s kompleksnim bolestima snažno podržavaju hipotezu da uz faktore poput epistaze i interakcija gena i okoliša, mnogi drugi genetički faktori značajnog učinka tek trebaju biti otkriveni kroz studije rijetkih varijanata. Podaci nastali u ovom projektu ne samo da pomažu interpretaciji svih genetičkih istraživanja nego daju primjer kako najbolje dizajnirati i analizirati istraživanja bolesti na temelju slijeda DNA.

Naposljetku, analiza varijanata niske frekvencije pokazala je posljedice selekcije na funkcionalno važnim mjestima u genomu i kako to može utjecati na populacijsku prošlost i dovesti do znatne lokalne diferencijacije, čak i kad su standardne mjere strukture kao što je F_{ST} vrlo niske. To se događa primarno zbog toga jer rijetki varijanti su recentnog podrijetla i geografski su vrlo ograničeno rasprostranjeni. Zbog toga interpretacija rijetkih varijanata neke jedinke s određenom bolešću bi trebala biti unutar konteksta lokalne, bilo geografske ili obiteljske, genetičke pozadine. Povrh toga, protivi se vrijednosti sekvenciranja jedinki iz različitih populacija kako bi se odredio spektar ljudske genetičke varijabilnosti i potkrjepilo istraživanje bolesti u više različitih populacija.

3.2.5. Druga faza

Tijekom druge faze podaci od preko 1700 jedinki su se koristili za razvoj metoda kako bi se poboljšale postojeće i razvile nove za analizu svojstava poput multialelnih varijanata (engl. multi allelic variant sites) i integraciju kompleksnih mutacija i strukturalnih varijanata.

3.2.6. Treće faza

U trećoj fazi genomi 2,504 osobe iz 26 populacija iz Afrike, Istočne i Južne Azije, Europe te Sjeverne i Južne Amerike su rekonstruirani kombinacijom sekvenciranja čitavog genoma niskom pokrivenošću (prosječna pokrivenost 4-7 X), eksoma visokom pokrivenošću (prosječna pokrivenost 65.7 X) i genotipizirani mikročipom (engl. dense microarray genotyping). Ukupno je pronađeno i opisano preko 88 milijuna varijanata - 84.7 milijuna pojedinačnih nukleotidnih polimorfizama, 3.6 milijuna kratkih insercija/delecija i 60,000 strukturalnih varijanata.

Pojedine jedinke i njihovi potomci su bili genotipizirani pomoću mikročipova (engl. high-density SNP microarrays). Tako se na jeftiniji način otkrila genetička varijabilnost i odredili pojedini genotipovi i haplotipovi. Više se nisu analizirale samo bi-alelne točkaste mutacije već i multi-alelne točkaste mutacije, inedeli i različiti setovi strukturalnih varijanata. Pri otkrivanju genetičkih varijanata korištena su 24 alata za analizu sljedova i klasifikatori strojnog učenja (engl. machine-learning classifiers) kako bi se visoko kvalitetni varijanti odvojili od mogućih netočnih pozitiva, te se time uravnotežila senzitivnost i specifičnost.

U ovoj zadnjoj fazi projekta stvoren je do sada najopširniji set ljudskih strukturalnih varijanti koji je vrijedan resurs za buduća istraživanja bolesti i populacijske genetike.

Procijenjeno je da svaka jedinka posjeduje prosječno 18.4 Mbp strukturalnih varijanata u diploidnom genomu, većinu njih čine multi-alelne kopije različitog broja ponavljanja (engl. multi-allelic copy number variation) (11.3 Mbp) i bi-alelne delecije (5.6 Mbp) (Sudmant et al., 2015).

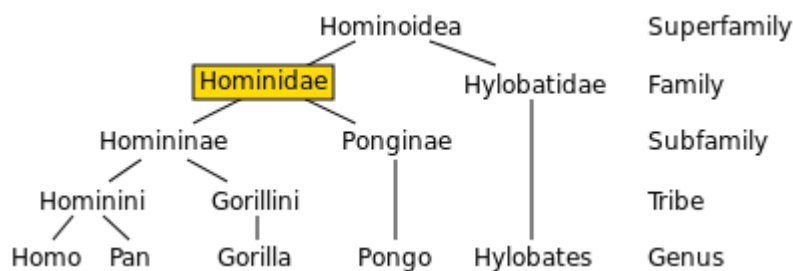
4. EVOLUCIJA I GENETIČKA VARIJABILNOST

Istraživanje ljudske genetičke varijabilnosti osim što ima medicinsku primjenu koristi se pri proučavanju i rekonstrukcije evolucijske povijesti čovjeka. Uzorci genetičke varijabilnosti informiraju nas o povijesti populacije zato što svaki veliki demografski događaj ostavlja trag na genetičkoj različitosti čitave populacije. Smanjenje veličine populacije reducira genetičku varijabilnost, a povećanje veličine populacije povećava varijabilnost. Izmjena jedinki između populacija neizbježno stvara veću genetičku sličnost dok izolacija pridonosi očuvanju genetičku jedinstvenost. Ovi demografski pokazatelji se prenose kroz generacije tako da genomi modernih jedinki reflektiraju njihovu demografsku prošlost. Akumulacijom podataka o genetičkoj varijabilnosti te razvojem alata za analizu, naša evolucijska povijest postaje sve jasnija.

Većina istraživanja evolucije čovjeka temelji se na analizi mitohondrijske DNA i nerekombinantnog (engl. non-recombining) dijela Y kromosoma. Budući da nisu podložni rekombinaciji, nasljeđuju se od samo jednog roditelja, prenose se relativno nepromijenjeni kroz generacije tako otkrivajući majčinsku i očinsku liniju populacije. Populacije dijele istu mitohondrijsku DNA ili Y kromosom kao rezultat zajedničkog podrijetla ili protoka gena.

Rezultate istraživanja na temelju mitohondrijske DNA i Y kromosoma treba uzimati s oprezom jer oboje predstavljaju jedan visoko varijabilan lokus, statističke procjene parametara koje se koriste nekada znaju biti krive. Puno pouzdanije procjene takvih parametara mogu se dobiti analizom više neovisnih lokusa (točkaste mutacije ili broj uzastopnih ponavljanja). Pri proučavanju ljudske evolucijske prošlosti genetički podaci se svako moraju kombinirati s arheoloških, lingvističkih i povijesnim podacima kako bi se što točnije interpretirala evolucija čovjeka.

4.1. Evolucija Hominida



Slika 4. Taksonomski odnosi nadporodice Hominoidea

Moderni čovjek *Homo sapiens* taksonomski pripada nadporodici *Hominoidea*, porodici *Hominidae*, podporodici Homininae, plemenu *Hominini* te rodu *Homo* (prema Slici 4.).

U porodicu *Hominidae* uz rod *Homo* ubrajaju se rodovi *Pan* (čimpanze), *Gorilla* (gorile) i *Pongo* (orangutani). Prema procjenama posljednji zajednički predak svih rodova porodice *Hominidae* živio je prije otprilike 14 milijuna godina, kad su se preci orangutana izdvojili od ostala tri roda.

Evolucija hominida je bilo pitanje koje je mučilo znanstvenike više od stoljeća međutim molekularni dokazi poput varijabilnosti genoma rodova porodice *Hominidae* uspjele su donekle odgovoriti na to pitanje.

Različiti dijelovi genoma imaju različitu razinu varijabilnosti između različitih rodova hominida. Dokazano je da divergentnost sljedova između ljudi i čimpanzi iznimno varira. Na primjer, divergentnost sljedova varira između 0% do 2.66 % unutar nekodirajućih, neponavljajućih genomskih regija ljudi i čimpanzi (Chen and Li, 2001).

Time i filogenetska stabla dobivena komparativnom analizom dijelova DNA ne odgovaraju uvijek filogenetskom stablu vrsta. Divergentnost sljedova iznimno varira između ljudi, čimpanzi i gorila. Prema većini DNA sljedova, ljudi i čimpanze su srodstveno bliži, međutim u nekim dijelovima genoma ljudi i gorile te čimpanze i gorile su sličniji.

Ljudski genom sadrži 23 para kromosoma dok genomi ostalih živućih hominida sadrže 24 para kromosoma. Ljudski kromosom 2 je nastao fuzijom dva kromosoma 2a i 2b koja su ostala odvojena kod ostalih primata.

4.1.1. Genetička varijabilnost hominida

Genom čovjeka i čimpanze razlikuje se za oko 35 milijuna točkastih mutacija te još oko 3% čitavog genoma razlikuje se zbog delecija, insercija i duplikacija (Chimpanzee Sequencing; Analysis Consortium, 2005.)

S obzirom da je stopa mutacija (engl. mutation rate) relativno konstantna, otprilike pola navedenih promjena pojavilo se u ljudskoj liniji. Samo mali dio tih razlika uzrokovalo je različit fenotip čovjeka i čimpanze, a pronalazak mutacija koje su to uzrokovale je iznimno teško. Većina mutacija su neutralne i ne utječu na fenotip. Molekularna evolucija možda djeluje na drugačije načine, kroz evoluciju proteina, gubitak gena, diferencijalnu gensku ekspresiju i RNA evoluciju.

Mnogo različitih mutacija može inaktivirati gen, ali samo neke mogu promijeniti funkciju gena na specifičan način. Inaktivacijske mutacije bi onda bile spremne za selekciju da djeluje na njih. Gubitak gena (engl. gene loss) bi mogao biti mehanizam evolucijske adaptacije (Olson, M. 1999). 80 gena je izgubljeno u ljudskoj liniji nakon odvajanja od zadnjeg zajedničkog pretka s čimpanzom od čega je 36 gena kodiralo za olfaktorne receptore. Drugo istraživanje govori da se izgubilo 86 gena (Demuth, J., Bie, T., Stajich, J., Cristianini, N. and Hahn, M., 2006).

Duplikacije su imale ulogu u stvaranju novih gena kod primata i oblikovanju ljudske genetičke varijabilnosti. Pri usporedbi ljudskog genoma s genomima čimpanze, gorile, orangutana te primata gibona i makaka, otkriveno je da samo čovjek sadrži otprilike 20,000 specifičnih insercija za koje se smatra da imaju regulatornu ulogu. Dok većina insercija je neutralna s obzirom na fitnes (engl. fitness), mali dio njih je identificiran u pozitivno selektiranim genima koji su povezani s živčanim sustavom i osjetilnom percepcijom. Ta otkrića potvrđuju važnost uloge insercija specifičnih za ljude u nedavnoj evoluciji čovjeka.

(Hellen, E. and Kern, A., 2015).

Ljudske ubrzane regije (engl. human accelerated regions) su dijelovi genoma koji se razlikuju između ljudi i čimpanzi do te mjere da se objašnjavaju genetičkim driftom. Ove regije pokazuju da su bile predmet prirodne selekcije, dovodeći do evolucije svojstava koji su specifični samo za ljude. Dva primjera su geni HAR1F i HAR2.

Gen HAR1F je povezan s razvojem mozga, a gen HAR2 je imao ulogu u razvoju odvojenog palca. Također, postoje hipoteze da je za razlike između ljudi i čimpanzi zaslužnija regulacija

genske ekspresije nego razlika u samim genima. Analize očuvanih nekodirajućih sljedova koji često sadrže funkcionalne i pozitivno selektirane regulatorne regije, potvrđuje ovu mogućnost (Bird et al., 2007).

Uspoređujući genome čovjeka i čimpanze, utvrđeno je da se razlikuju 1.23 % zbog pojedinačnih supstitucija baza. Od tih supstitucija, za 1,06 % ili manje se smatra da predstavljaju fiksne razlike između te dvije vrste dok ostatak su varijante u čovjeku ili čimpanzi. Indeli (insercije/ delecije) čine manji broj razlika, ali doprinose ~1.5% unikatnim sljedovima svakog genoma, budući da insercije ili delecije mogu obuhvatiti od jedne do milijun baza (Chimpanzee Sequencing; Analysis Consortium, 2005.).

Također su proučavane i uspoređivane segmentalne duplikacije (engl. segmental duplications) u oba genoma, čije insercije i delecije u genomu obuhvaćaju mnogo indel sljedova. Otkriveno je da ukupno 2.7 % eukromatinskih sljedova bilo djelomično duplicirano u jednoj ili drugoj liniji (Cheng et al., 2005).

Divergentnost sljedova u hominida ima sljedeći uzorak: čovjek-čimpanza < čovjek-gorila << čovjek-orangutan.

4.2. Evolucija čovjeka

Glavno pitanje evolucije čovjeka je podrijetlo modernog čovjeka. Kao odgovor na to pitanje predložene su dvije suprotne teorije.

Prva je Multiregionalna teorija koja tvrdi da je moderan čovjek nastao iz više arhaičnih oblika poput Neandertalca u različitim dijelovima Starog Svijeta (Afrika, Azija i Europa). Genetičke sličnosti uočene u sadašnjim populacijama su rezultat protoka gena između tih populacija i prirodne selekcije za slična adaptivna svojstva.

Prema suprotnoj teoriji, nazvanoj “Iz Afrike hipoteza” (engl. Out of Africa hypothesis) koja govori da je moderan čovjek nastao najprije u Africi te potom migrirao Europu i Aziju, prije manje od 100,000 godina, kako bi zamijenio postojeće arhaične populacije.

Većina genetičkih dokaza potkrepljuje teoriju da je moderan čovjek prvo nastao u Africi. Prema gotovo svim genetičkim istraživanjima najveća varijabilnost genoma nalazi se u afričkim populacijama što ide u prilog teoriji o podrijetlu iz Afrike. Varijabilnost je funkcija vremena i veličine populacije, makar to može biti i rezultat samo nekadašnje velike populacije u Africi. Činjenica je da većina genetičke varijabilnosti u ne-afričkim populacijama samo dio onoga što je pronađeno u afričkim populacijama koje su puno bogatije različitim genetičkim varijantama te se to ne može objasniti Multiregionalnom teorijom. Ne-afričke populacije su puno sličnije jedna drugoj nego što su slične afričkim što također navodi da je dio afričke populacije napustio kontinent i kolonizirao ostatak svijeta. Genetička otkrića također navode da je rana ljudska populacija bila vrlo mala jedno vrijeme (otprilike 10,000 jedinki). To opovrgava Multiregionalnu teoriju zato jer za kontinuirani protok gena između ljudskih populacija koje su razbacane Starim Svijetom treba puno veća populacija jedinka. Iako većina dokaza ide u prilog podrijetlu modernog čovjeka iz Afrike, teško je za prihvatiti da nekadašnje arhaične populacije iz drugih dijelova svijeta nisu genetički doprinjele današnjoj populaciji, tj. da nije došlo do miješanja između anatomski modernog čovjeka iz Afrike i arhaičnih populacija iz Europe ili Azije.

Slika o evoluciji čovjeka bit će pun jasnija kad autosomalni sljedovi budu analizirani i uspoređeni između više ljudskih populacija.

Procjenom stope mutacija u mitohondrijskoj DNA, dob zajedničke ancestralne mitohondrijske DNA može biti određena: zajednička ancestralna DNA povezuje mitohondrijske DNA koje su divergirale prosječno 0.57 %. Pretpostavljajući da stopa

mutacije iznosi 2% do 4% u milijun godina, to podrazumijeva da je zajednički predak svih današnjih mitohondrijskih DNA postojao prije 140,000 - 290,000 godina (Cann, RL; Stoneking, M; Wilson, AC, 1987.). Prema mnogim znanstvenicima, ova procjena je vrlo gruba.

Y kromosom je mnogo veći od mitohondrijske DNA i relativno je homogen te sadrži više genetičke informacije. Istraživanja Y kromosoma pokazuju slične rezultate kao i ona s mitohondrijskom DNA. Procjena dobi postojanja ancestralnog Y kromosoma za sve postojeće Y kromosome je otprilike prije 70,000 godina. Prema genetičkoj varijabilnosti, Y kromosoma potječe iz Afrike i potomci izvedene linije su napustili Afriku i s vremenom su bili zamijenjeni arhaičnim ljudskim Y kromosomom u Euroaziji (Henn et al., 2012).

Podrijetlo modernog čovjeka iz Afrike potvrđuje i činjenica da promjene u veličini lubanje se smanjuju s udaljenošću iz Afrike istom brzinom kako se smanjuje genetička raznolikost. Ljudska genetička varijabilnost manja je u populacijama s većom migracijskom udaljenošću od Afrike te se smatra da j to zbog efekta uskog grla boce (engl. bottle neck effect) tijekom ljudskih migracija, a to je događaj kada se privremeno smanji veličina populacije (Manica et al., 2007).

Procjenjuje se da moderan čovjek vjerojatno potječe s područja Namibije i južne Afrike, i da je dio njih napustio Afriku kroz istok Afrike. U prilog tome ide činjenica da afričke populacije imaju najviše genetičke varijabilnosti nego bilo gdje na Zemlji i da genetička struktura Afrikanaca vodi u trag do 14 ancestralnih populacijskih grupa koje koreliraju s nacionalnošću i kulturom ili jezikom. Istraživanje koje to potvrđuje je trajalo deset godina i analiziralo varijacije 1,327 DNA markera iz 121 afričke populacije, 4 afričko-američke populacije i 60 ne-afričkih populacija.

4.2.1. Populacijska genetika

Smatra se da se na temelju distribucije ljudske genetičke varijabilnosti među modernim ljudima može odrediti ljudska demografska povijest. Dalje se raspravlja jesu li ljudi prošli kroz efekt uskog grla boce prije nagle ekspanzije koja se preklapa s migracijom iz Afrike dovodeći do afričko-euroazijske divergencije prije 100,000 godina, nakon koje je slijedila europsko-azijska divergencija prije oko 40,000 godina. Kad su prvi moderni ljudi napustili Afriku i kolonizirali ostatak svijeta i dalje je predmet niza rasprava.

Međutim, nagla ekspanzija nekada male populacije ima dvije važne posljedice na distribuciju genetičke varijabilnosti. Prvo, takozvani efekt osnivača (engl. founder effect) se pojavljuje kada osnivačke populacije donose samo dio genetičke varijabilnosti iz svojih predačkih populacija. Drugo, kako osnivači postaju više geografski odvojeni, vjerojatnost da dvije jedinke iz različitih osnivačkih populacija se razmnožavaju postaje manja. Posljedica ovakvog razmnožavanja je smanjenje razmjene gena između geografskih grupa i povećanje genetičke udaljenosti između njih.

Ekspanzija ljudi iz Afrike utjecala je na distribuciju genetičke varijabilnost na dva načina. Manje populacije imaju veći genetički drift zbog povećanih fluktuacija neutralnih polimorfizama. Novi polimorfizmi koji se pojavljuju u jednoj grupi će se manje vjerojatno prenijeti u druge grupe jer je protok gena smanjen.

Populacije u Africi osim što imaju veću genetičku varijabilnost imaju i manje povezanosti neravnoteže (engl. linkage disequilibrium) nego populacije izvan Afrike, djelomično zato jer su populacije ljudi u Africi bile veće tijekom povijesti i zato jer broj ljudi koji je napustio Afriku da kolonizirao ostatak svijeta bio relativno malen (Gabriel *et al.* 2002).

4.2.2. Križanje modernog čovjeka s arhaičnim vrstama

Postoje hipoteze da se moderan čovjek križao s Neandertalcem tijekom srednjeg paleolitika. Sekvenciranjem genoma Neandertalca i usporedbom s genomom modernog čovjeka dokazalo se da postoji mogućnost da je do toga došlo jer mali dio genoma Neandertalca je prisutan u genomu modernog čovjeka iz euroazijskih populacija, ali nije prisutan u genomu afričkih populacija. Najzanimljivija otkrića usporedbe genoma Neandertalca i modernog čovjeka je da oboje dijele varijantu gena laktaze koja uzrokuje netoleranciju na laktozu tako što kodira za enzim koji ne može razgraditi laktozu iz mlijeka. Osim te varijante dijele i varijantu gena FOXP2 koji je povezan s razvojem mozga i govora, navodeći mogućnost da je Neandertalac mogao govoriti (Dumas et al., 2007).

Smatra se da između 4 % i 6 % genoma Melanezijaca je podrijetlom od Denisovskog čovjeka- nekadašnje nepoznate vrste koja dijeli podrijetlo s Neandertalcem. Vjerojatno je do križanja došlo tijekom rane migracije predaka Melanezijaca u sjeverno-istočnu Aziju dok je Denisovski čovjek bili prošireni čitavom istočnom Azijom (Reich et al., 2010).

Tako da Melanezijci su populacija s najviše primjesa arhaičnih populacija, imaju čak ~8% genoma podrijetlom od Denisovskog čovjeka/Neandertalaca.

Prema posljednjim istraživanjima na temelju čitavih genoma uočeno je da je introgresija veća u azijskim populacijama nego europskim (Wall et al., 2013). Također u genomima afričkih populacija pronađeni su tragovi protoka gena s arhaičnim ljudskim precima koji navode na određena niska razina protoka gena je bila prisutna u vremenu i prostoru tijekom evolucije anatomske modernog čovjeka (Hammer et al., 2011).

5. LJUDSKA GENETIČKA VARIJABILNOST I RASE

Novonastali podaci o ljudskoj genetičkoj varijabilnosti su pokrenuli raspravu postoji li biološka podloga za kategorizaciju ljudi u rase. Čovjek je evolucijski vrlo mlad organizam i kao vrsta je vrlo homogen o čemu govori veći prije spomenuti podataka da svi ljudi dijele oko 99.5 % sličnosti u slijedu DNA. Iako su razlike između ljudskih populacija male, postoje razlike u određenim genima kao što su geni ABCC11 i SLC24A5 koji su nazvani markerima podrijetla (engl. ancestry-informative markers) te služe kako bi se jedinke smjestile u široke, geografski bazirane grupe.

Kompjuterska analiza stotine polimorfnih lokusa iz populacija koje su rasprostranjene diljem svijeta otkrila je postojanja genetičkih grupacija koja su grubo povezana s grupama koje su povijesno zauzimale veće kontinentalne i subkontinentalne regije (Rosenberg *et al.* 2002; Bamshad *et al.* 2003).

Risch i suradnici tvrde da takvi uzorci varijabilnost daju biološko opravdanje za upotrebu tradicionalnih rasnih kategorija. Takvo kontinentalno grupiranje poklapa se s podjelom ljudi na sub-saharske Afrikance, Europljane, zapadne Azijate, centralne Azijate, južne Azijate, sjeverne Amerikance, istočne Azijate itd. (Risch *et al.* 2002). Dok s druge strane neki zastupaju tezu da isti podaci ruše tradicionalne podjele na rasne grupe (King and Motulsky 2002; Calafell 2003; Tishkoff and Kidd 2004). Oni tvrde da na primjer velike populacije smatrane rasama ili subgrupama unutar rasa ne formiraju nužno vlastiti genetičku grupu.

Nadalje, kako je većina ljudske genetičke varijabilnosti gradualna, mnoge jedinke potječu iz dvije ili više kontinentalnih grupa. Tako da genetički utemeljeno “biogeografsko podrijetlo” neke jedinke je zapravo vrlo široko i popraćeno određenim nesigurnostima (Pfaff *et al.* 2004). U mnogim dijelovima svijeta, populacije su se miješale na takav način da mnoge jedinke imaju relativno nedavne pretke iz odvojenih regija. Makar genetička analiza velikog broja lokusa može stvoriti procjene udjela pojedinih predaka koji dolaze iz različitih kontinentalnih regija (Bamshad *et al.* 2004), iako te procjene mogu biti netočne jer ljudske populacije izmjenjuju partnere od lokalne do kontinentalne skale kroz čitavu povijest (Chisholm *et al.*, 1995). S velikim brojem markera, informacije za procjenu razine miješanja jedinki ili grupa je ograničena i procjene obično sadrže široke intervale raspona (Pfaff *et al.* 2004).

6. SAŽETAK

Ljudska genetička varijabilnost predstavlja genetičke razlike unutar i između populacija ljudi. Razvojem tehnologija sekvenciranja genoma i metoda analize sljedova omogućeno je istraživanje genetičke varijabilnosti ljudi kao nikada do sada. Postoji više oblika genetičke varijabilnosti, ali glavna je podjela na pojedinačne nukleotidne polimorfizme i strukturalne varijante. Glavna primjena genetičke varijabilnosti je u medicini kako bi se određeni genotipovi povezali s kompleksnim bolestima te u istraživanju evolucije čovjeka.

SUMMARY

Human genetic variation is the genetic differences both within and among populations. The development of the genome sequencing and the methods of sequence analysis have made the investigation of human genetic variation more accurate than ever before. There are multiple types of genetic variation but most common classification contains single nucleotide polymorphisms and structural variants. The study of human genetic variation has both evolutionary significance and medical applications.

7. LITERATURA

Altshuler, D., Gibbs, R., Peltonen, L., Altshuler, D., Gibbs, R. et.al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), pp.52-58.

Auton, A., Abecasis, G., Altshuler, D., Durbin, R., Abecasis, G., et.al. (2015). A global reference for human genetic variation. *Nature*, 526(7571), pp.68-74.

Bamshad, M., Wooding, S., Salisbury, B. and Stephens, J. (2004). Deconstructing the relationship between genetics and race. *Nat Rev Genet*, 5(8), pp.598-609.

Bird, C., Stranger, B., Liu, M., Thomas, D., Ingle, C., Beazley, C., Miller, W., Hurles, M. and Dermitzakis, E. (2007). Fast-evolving noncoding sequences in the human genome. *Genome Biol*, 8(6), p.R118.

Cann, RL; Stoneking, M; Wilson, AC (1987). "Mitochondrial DNA and human evolution". *Nature*. **325** (6099): 31–6.

Chen, F. and Li, W. (2001). Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees. *The American Journal of Human Genetics*, 68(2), pp.444-456.

Chimpanzee Sequencing; Analysis Consortium (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome". *Nature*. **437** (7055): 69–87. Bibcode:2005Natur.437...69.. doi:10.1038/nature04072. PMID 16136131.

Chisholm, B., Cavalli-Sforza, L., Menozzi, P. and Piazza, A. (1995). The History and Geography of Human Genes. *The Journal of Asian Studies*, 54(2), p.490.

Collins, F. S.; Brooks, L. D.; Chakravarti, A. (1998). "A DNA polymorphism discovery resource for research on human genetic variation". *Genome Research*. **8** (12): 1229–1231. PMID [9872978](#)

Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., Church, D., DeJong, P., Wilson, R., Pääbo, S., Rocchi, M. and Eichler, E. (2005). A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature*, 437(7055), pp.88-93.

Demuth, J., Bie, T., Stajich, J., Cristianini, N. and Hahn, M. (2006). The Evolution of Mammalian Gene Families. *PLoS ONE*, 1(1), p.e85.

Durbin, R., Altshuler, D., Durbin, R., Abecasis, G., Bentley, D., Chakravarti, et.al. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), pp.1061-1073.

Frazer, K., Murray, S., Schork, N. and Topol, E. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet*, 10(4), pp.241-251.

Hammer, M., Woerner, A., Mendez, F., Watkins, J. and Wall, J. (2011). Genetic evidence for archaic admixture in Africa. *Proceedings of the National Academy of Sciences*, 108(37), pp.15123-15128.

Hellen, E. and Kern, A. (2015). The Role of DNA Insertions in Phenotypic Differentiation between Humans and Other Primates. *Genome Biology and Evolution*, 7(4), pp.1168-1178.

Henn, B., Botigué, L., Gravel, S., Wang, W., Brisbin, A., Byrnes, J., Fadhlouzi-Zid, K., Zalloua, P., Moreno-Estrada, A., Bertranpetit, J., Bustamante, C. and Comas, D. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genetics*, 8(1), p.e1002397.

Jorde, L. and Wooding, S. (2004). Genetic variation, classification and 'race'. *Nature Genetics*, 36(11s), pp.S28-S33.

Levy, S., Sutton, G., Ng, P., Feuk, L., Halpern, A., Walenz, B., Axelrod, N., Huang, J., Kirkness, E., Denisov, G., Lin, Y., MacDonald, J., Pang, A., Shago, M., Stockwell, T., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S., Busam, D., Beeson, K., McIntosh, T., Remington, K., Abril, J., Gill, J., Borman, J., Rogers, Y., Frazier, M., Scherer, S., Strausberg, R. and Venter, J. (2007). The Diploid Genome Sequence of an Individual Human. *PLoS Biology*, 5(10), p.e254.

Manica, A., Amos, W., Balloux, F. and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. *Nature*, 448(7151), pp.346-348.

McVean, G., Altshuler (Co-Chair), D., Durbin (Co-Chair), R., Abecasis, G., Bentley, D., et.al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), pp.56-65.

Ng, P., Levy, S., Huang, J., Stockwell, T., Walenz, B., Li, K., Axelrod, N., Busam, D., Strausberg, R. and Venter, J. (2008). Genetic Variation in an Individual Human Exome. *PLoS Genetics*, 4(8), p.e1000160.

Olson, M. (1999). When Less Is More: Gene Loss as an Engine of Evolutionary Change. *The American Journal of Human Genetics*, 64(1), pp.18-23.

Reich, D., Green, R., Kircher, M., Krause, J., Patterson, N., Durand, E., Viola, B., Briggs, A., et.al. *Nature*, 468(7327), pp.1053-1060.

Rosenberg, N. (2002). Genetic Structure of Human Populations. *Science*, 298(5602), pp.2381-2385.

Sudmant, P., Rausch, T., Gardner, E., Handsaker, R., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., et. al.(2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), pp.75-81.

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S., Zhu, X., Brown, A., Pankow, J., Province, M., Hunt, S., Boerwinkle, E., Schork, N. and Risch, N. (2005). Genetic Structure, Self-Identified Race/Ethnicity, and Confounding in Case-Control Association Studies. *The American Journal of Human Genetics*, 76(2), pp.268-275.

Tishkoff, S. and Kidd, K. (2004). Implications of biogeography of human populations for 'race' and medicine. *Nature Genetics*, 36(11s), pp.S21-S27.

Wall, J., Yang, M., Jay, F., Kim, S., Durand, E., Stevison, L., Gignoux, C., Woerner, A., Hammer, M. and Slatkin, M. (2013). Higher Levels of Neanderthal Ancestry in East Asians than in Europeans. *Genetics*, 194(1), pp.199-209.

Wang, X., Grus, W. and Zhang, J. (2006). Gene Losses during Human Origins. *PLoS Biology*, 4(3), p.e52.